Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 71 (2005) 1–13
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

On the distribution of the clipping median under a mixture model

Ansgar Steland*

Fakultät für Mathematik, Ruhr-Universität Bochum, 3 NA 3/71, Universitätsstr. 150, D-44780 Bochum, Germany

Received 7 February 2003; received in revised form 6 October 2004
Available online 13 November 2004

Abstract

Motivated by applications in image processing, quality control, and econometrics we derive the exact distribution function of the clipping median estimator which is designed to provide simultaneously robust smooth and jump-preserving reconstructions. We allow for a mixture model which is of special interest for applications in pixel-wise object detection. To construct statistical tests for pixel classification, we propose to rely on estimated p-values. Simulations suggest that the resulting approximations are reliable.

Keywords: Change point; Econometrics; Nonparametric regression; Image processing; Jump-preserving estimation; Edge detection; Robust statistics

1. Introduction

Suppose we are given n real-valued random variables (r.v.'s) $Y_1, \ldots, Y_n$ and aim at testing whether $m_n = E(Y_n)$ is affected by a level shift when compared to $m_1, \ldots, m_{n-1}$, where $m_i = E(Y_i)$, $i = 1, \ldots, n$. This problem arises in various contexts, e.g., when a sequential monitoring scheme for an economic time series or a series of quality measurements is established for the first time and one wants to compare the current observation Y_n with past data $Y_1, \ldots, Y_{n-1}$ to detect a level shift with no delay. Such a comparison can also be an useful additional tool for a fixed

*Fax: +49 234 321 4559.
E-mail address: ansgar.steland@ruhr-uni-bochum.de (A. Steland).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2004.10.016

sample analysis, if the analysis is naively applied in a sequential way after getting the new observation Y_n, as it is often done in empirical econometric research. In image processing an important task is to decide whether a pixel belongs to the background or to the foreground (object). In this case we may define a local neighbourhood of $h = n - 1$ neighbouring pixels, denote their grey values by $Y_1, \ldots, Y_{n-1}$, and add the grey value Y_n of the current pixel to the sample. In this article we focus on this image processing problem, last but not least because the estimator has its origins in that field, but our results can be applied in various settings.

Quite often primary interest is in detecting level jumps. An estimator which is especially designed to reproduce (grey level) jumps defining edges and contour lines of objects and their location with high precision is the *clipping median* of m_n defined as the empirical median of all Y_i's with similar values as Y_n. Here we consider a slightly more general version. Let $k \geqslant 0$ be a symmetric function $\mathbb{R} \to \mathbb{R}^+$, usually but not necessarily unimodal and decreasing, and $M > 0$ a parameter. Common choices for k are the uniform kernel or the Gaussian kernel. Define the clipping median estimator by

$$\widehat{m}_n = \mathrm{ClipMed}\{k([Y_i - Y_n]/M)Y_i\},$$

where ClipMed is defined as

$$\mathrm{Med}\{k([Y_i - Y_n]/M)Y_i : 1 \leqslant i \leqslant n \text{ with } |Y_i - Y_n| \leqslant M\}. \tag{1}$$

We use the common definition

$$\mathrm{Med}\{\xi_1, \ldots, \xi_n\} = \begin{cases} \frac{1}{2}(\xi_{(n/2+1)} + \xi_{(n/2)}) & n \text{ even}, \\ \xi_{([n+1]/2)} & n \text{ odd} \end{cases}$$

of the empirical median of n r.v.'s $\xi_1, \ldots, \xi_n$, where $\xi_{(1)} \leqslant \cdots \leqslant \xi_{(n)}$ denotes the corresponding order statistic. Recall that the exact distribution for i.i.d. observations is related to the binomial distribution due to the relationship

$$\mathrm{Med}\{\xi_1, \ldots, \xi_n\} \leqslant x \Leftrightarrow \sum_{i=1}^{n} \mathbf{1}(\xi_i \geqslant x) \leqslant (n+1)/2,$$

which holds true for both even and odd n. For the clipping median the situation is more involved and is the topic of this paper.

The particular version (1) has been studied in Pawlak et al. (2004), where sufficient conditions were derived which ensure that the sequential stopping rule $\inf\{n \in \mathbb{N} : \widehat{m}_n > c\}$ can detect jumps in time series with no delay, if c is appropriately chosen. Note that $\widehat{m}_n$ reduces to the empirical median of all Y_i such that $|Y_i - Y_n| \leqslant M$ if k is the uniform kernel. However, the behaviour of $\widehat{m}_n$ for general k is as follows. Observations with $|Y_i - Y_n| > M$, i.e., far away from the current observation, are excluded from the calculation (clipping). The kernel k performs a shrinkage operation, since observations Y_i with large values of $|Y_i - Y_n|$, which are not ignored by the clipping mechanism, are shrunken towards 0. This means, the data transformation $Y_i \mapsto Z_i = k([Y_i - Y_n]/M)Y_i$ forms a cluster. The clipping mechanism ensures that the sample from which the median is calculated shrinks substantially, if the current observation is different from the other ones. This property ensures that sufficiently large level shifts can be reproduced more accurately than with classic averaging procedures. If $|Y_i - Y_n| > M$ for all $1 \leqslant i < n$, the clipping median interpolates, i.e., $\widehat{m}_n = Y_n$. Finally note that if the sample is homogenous (i.i.d.) with median 0, we have $\mathrm{Med}k([Y_i - Y_n]/M)Y_i = 0$, $1 \leqslant i \leqslant n$. If there is a level shift, the Z_i's before the shift will be shrunken to 0 if k is chosen appropriately, whereas the Z_i's after the shift are not shrunken.

The basic idea underlying the estimator $\widehat{m}_n$ can be traced back to Lee (1983) who studied the problem of edge-preserving estimators for image processing purposes. Lee's estimator is implemented in many image processing packages including Mathematica, see Wolfram Research (2004). It has been studied in different contexts, e.g., to smooth magnetic resonance (MR) images. That application has been studied by Godtliebsen (1991), Godtliebsen and Spjøtvoll (1991), Chapter 4 of Budinger et al. (1996), and Chiu et al. (1998). The latter paper also discusses the relationship to M estimation. For a discussion of an application of nonlinear Gaussian filters to images we refer to Godtliebsen and Marron (1997). Further recent work can be found in Pawlak and Rafajłowicz (2001). The application of jump-preserving estimators for sequential monitoring and related theoretical results can be found in Pawlak and Rafajłowicz (2000), Steland (2002a), Steland (2004a), and Pawlak et al. (2004). For related recent results on classic kernel estimators we refer to Steland (2004b) and the references given there.

The contribution of this article is to provide a theoretical basis for deriving statistical test procedures based on the clipping median. We consider two models. In Section 2 we study the d.f. of $\widehat{m}_n$ assuming that the neighbourhood $Y_1, \ldots, Y_{n-1}$ forms a homogenous i.i.d. sample. Section 3 considers a more general mixture model. Finally, in Section 4 we study the a.s. convergence of the proposed estimated p-values.

2. i.i.d. Neighbourhoods

Let us first study a setting which can be used to decide whether or not a single pixel belongs to a homogeneous background. Assume $Y_1, \ldots, Y_n$ are independent such that $Y_1, \ldots, Y_{n-1}$ are i.i.d. with common d.f. $F(x)$, and $Y_n \sim F(x - \Delta)$ for some shift $\Delta \in \mathbb{R}$. We aim at testing $\mathrm{H}_0 : \Delta = 0$ versus $\mathrm{H}_1 : \Delta \neq 0$ based on the test statistic $\widehat{m}_n$. Let

$$G_n(x; \Delta) = P_\Delta(\widehat{m}_n \leqslant x), \quad x \in \mathbb{R},$$

denote the d.f. of the clipping median estimator $\widehat{m}_n$.

Further, let $p_1, p_2 \in [0, 1]$ be two probabilities and denote by

$$q(n, i, k, p_1, p_2) = \frac{n!}{i!k!(n-i-k)!} p_1^i p_2^k [1 - p_1 - p_2]^{n-i-k} \tag{2}$$

the trinomial probabilities, where $n \in \mathbb{N}$, $i, k \geqslant 0$, and $i + k \leqslant n$. We set $q(n, i, k) = 0$ whenever the constraints on the indices are not satisfied. For an event A (or logical expression) $\mathbf{1}(A)$ is 1 if A occurs (is true), and equals 0 otherwise.

Theorem 1. *Let $Y_1, \ldots, Y_n$ be independent random variables such that $Y_1, \ldots, Y_{n-1}$ are i.i.d. with c.d.f. F and $Y_n \sim F(x - \Delta)$ for some shift $\Delta \in \mathbb{R}$. For all $x \in \mathbb{R}$ we have*

$$G_n(x; \Delta) = \int \sum_{(k,i) \in \mathscr{I}(x,y)} q(n-1, k, i, p_+(x, y), p_-(x, y)) \, \mathrm{d}F(y - \Delta),$$

where $\mathscr{J} = \{(k,i) \in \{0,\ldots,n-1\}^2 : k-i \leqslant 2\mathbf{1}(y > k(0)/M)\}$ *and*

$$p_+(x,y) = P(|Y_1 - y| \leqslant M, k([Y_1-y]/M)Y_1 \geqslant x),$$
$$p_-(x,y) = P(|Y_1 - y| \leqslant M, k([Y_1-y]/M)Y_1 < x)$$

for $x, y \in \mathbb{R}$.

Proof. We shall consider the case $\Delta = 0$, the general case $\Delta \in \mathbb{R}$ will be obvious. Hence, we have for each $x \in \mathbb{R}$

$$G_n(x;0) = \int P[\mathrm{Med}\{k([Y_i - y]/M)Y_i : |Y_i - y| \leqslant M\} \leqslant x | Y_n = y]\,\mathrm{d}F(y).$$

To calculate the integrand let

$$\mathscr{L}(y) = \{i \in \{1,\ldots,n\} : |Y_i - y| \leqslant M\}$$

be the random set of all indices corresponding to the clipped observations and put $L(y) = |\mathscr{L}(y)|$. Note that $n \in \mathscr{L}(y)$. Recall that by definition of the clipping median,

$$\mathrm{ClipMed}\{Z_i(y)\} \leqslant x \;\Leftrightarrow\; S_{L(y)}(x) \leqslant [L(y)+1]/2$$

for both $L(y)$ odd and even, where

$$S_l = \sum_{i=1}^{l} \mathbf{1}(Z_i(y) \geqslant x), \quad l \in \mathbb{N}$$

with $Z_i(y) = k([Y_i - y]/M)Y_i$, $i = 1,\ldots,n$. Consequently,

$$G_n(x;0) = \int P[S_{L(y)}(x) - (L(y)+1)/2 \leqslant 0 | Y_n = y]\,\mathrm{d}F(y).$$

We have

$$L(y) = \sum_{i=1}^{n-1} \mathbf{1}(|Y_i - y| \leqslant M) + 1,$$

$$\begin{aligned} S_{L(y)}(x) &= \sum_{i \in \mathscr{L}(y)} \mathbf{1}(Z_i(y) \geqslant x) \\ &= \sum_{i=1}^{n-1} \mathbf{1}(|Y_i - y| \leqslant M)\mathbf{1}(Z_i(y) \geqslant x) + \mathbf{1}(y \geqslant x/k(0)). \end{aligned}$$

Thus, $S_{L(y)}(x) - L(y)/2 \leqslant 0$ is equivalent to

$$\sum_{i=1}^{n-1} \eta_i(x,y) \leqslant 2\mathbf{1}(y > k(0)/M),$$

where

$$\eta_i(x,y) = 2\mathbf{1}(|Y_i - y| \leqslant M)\{\mathbf{1}(Z_i(y) \geqslant x) - 1/2\} \quad i = 1,\ldots,n-1$$

are i.i.d. $\{-1, 0, +1\}$-valued r.v.'s with

$$P(\eta_1(x,y) = k \mid Y_n = y) = p_-(x,y)^{\mathbf{1}(k=-1)} p_+(x,y)^{\mathbf{1}(k=1)} [1 - p_+(x,y) - p_-(x,y)]^{\mathbf{1}(k=0)}.$$

By independence of $\eta_i(x,y)$, $i = 1, \ldots, n$, the random vector $(N_+(x,y), N_-(x,y))$, where

$$N_+(x,y) = \sum_{i=1}^{n-1} \mathbf{1}(\delta_i(x,y) = +1),$$

$$N_-(x,y) = \sum_{i=1}^{n-1} \mathbf{1}(\delta_i(x,y) = -1)$$

follows a trinomial distribution with parameters $n-1, p_+(x,y)$, and $p_-(x,y)$, Therefore, we obtain

$$\begin{aligned} &P[S_{L(y)} - (L(y)+1)/2 \leqslant 0 \mid Y_n = y] \\ &\qquad = P[N_+(x,y) - N_-(x,y) \leqslant 2\mathbf{1}(y > k(0)/M) \mid Y_n = y] \\ &\qquad = \sum_{(k,i) \in \mathscr{I}(x,y)} q(n-1, k, i, p_+(x,y), p_-(x,y)). \end{aligned}$$

Obviously, the last expression is integrable w.r.t. $\mathrm{d}F(y)$. Hence,

$$G_n(x; \Delta) = \int \sum_{(k,i) \in \mathscr{I}(x,y)} q(n-1, k, i, p_+(x,y), p_-(x,y))\, \mathrm{d}F(y).$$

If $Y_n \sim F(x - \Delta)$, the measure $\mathrm{d}F(y)$ is replaced by $\mathrm{d}F(y - \Delta)$. $\square$

This result suggests the following estimators for the d.f. of the clipping median under both the null hypothesis and the alternative. For $\Delta = 0$ let

$$\begin{aligned} \widehat{G}_{L_1,L_2,n}(x;0) &= \int \sum_{(k,i) \in \mathscr{I}^*(x,y)} q(n-1, k, i, p_+(x,y), p_-(x,y)\, \mathrm{d}\widehat{F}_n(y) \\ &= \frac{1}{n-1} \sum_{(k,i) \in \mathscr{I}^*(x,Y_j)} q(n-1, k, i, p_+(x, Y_j), p_-(x, Y_j), \end{aligned} \tag{3}$$

where

$$\begin{aligned} \mathscr{I}^*(x,y) = \{&(k,i) \in \{0, \ldots, n-1\}^2 : |k - (n-1)p_+(x,y)| \leqslant L_1, \\ &|i - (n-1)p_-(x,y)| \leqslant L_2, |k - i| \leqslant 2\mathbf{1}(y > k(0)/M)\}. \end{aligned}$$

L_1, L_2 are truncation constants selecting the central atoms around the means. $\widehat{F}_{n-1}(x) = (n-1)^{-1} \sum_{j=0}^{n-1} \mathbf{1}(Y_j \leqslant x)$ is the e.d.f. of $Y_1, \ldots, Y_{n-1}$, and

$$\widehat{p}_1(x,y) = \frac{1}{n-1} \sum_{j=1}^{n-1} \mathbf{1}(|Y_{n-j} - y| \leqslant M, Z_{n-j}(y) \geqslant x), \tag{4}$$

$$\widehat{p}_2(x,y) = \frac{1}{n-1} \sum_{j=1}^{n-1} \mathbf{1}(|Y_{n-j} - y| \leqslant M, Z_{n-j}(y) < x). \tag{5}$$

For $\varDelta \neq 0$ the d.f. can be estimated by replacing $\mathrm{d}\widehat{F}_{n-1}(y)$ in (3) by $\mathrm{d}\widehat{F}_{n-1}(y-\varDelta)$. Thus, define $\widehat{G}_n(x;\varDelta)$ as

$$\frac{1}{n-1}\sum_{(k,i)\in\mathscr{J}^*(x,Y_j+\varDelta)} q(n-1,k,i,p_+(x,Y_j+\varDelta),p_-(x,Y_j+\varDelta)), \tag{6}$$

where $\widehat{p}_+(x,y)$ and $\widehat{p}_-(x,y)$ are defined as in (4) and (5), respectively.

We now propose to test $\mathrm{H}_0:\varDelta=0$ against $\mathrm{H}_1:\varDelta\neq 0$ using the test ϕ given by

$$\phi = \mathbf{1}(\widehat{p}_{L_1,L_2,n} \notin (\alpha/2, 1-\alpha/2)), \tag{7}$$

where

$$\widehat{p}_{L_1,L_2,n} = \widehat{G}_{L_1,L_2,n}(\widehat{m}_n;0)$$

is the estimated p-value. The a.s. convergence of the estimated p-value will be discussed in Section 4 using the more general framework of the next section.

3. Mixture model for neighbourhoods

We shall now generalize the results of the previous section to a mixture model for the neighbourhood of the current observation Y_n. So let $Y_1,\ldots,Y_n$ be independent r.v.'s. We assume that k Y_i's are distributed according to $F(x)$ (background), $n-1-k$ have distribution $F(x-\varDelta)$ (object), and $Y_n \sim F(x-\varDelta')$. This means, there exists a decomposition $\{1,\ldots,n-1\} = \mathscr{I}_1 + \mathscr{I}_2$ into disjoint nonempty sets $\mathscr{I}_1 \subset \{1,\ldots,n-1\}$ and $\mathscr{I}_2 \subset \{1,\ldots,n-1\}$ with $|\mathscr{I}_1| = k$ and $|\mathscr{I}_2| = n-1-k$, such that

$$Y_i \sim F(x), \quad i \in \mathscr{I}_1, \tag{8}$$

$$Y_i \sim F(x-\varDelta), \quad i \in \mathscr{I}_2. \tag{9}$$

Further, assume

$$Y_n \sim F(x-\varDelta'). \tag{10}$$

Here $\varDelta,\varDelta' \in \mathbb{R}$ are level shifts. Provided $\varDelta$ is known, the testing problem of interest is $\mathrm{H}_0:\varDelta'=\varDelta$ versus $\mathrm{H}_1:\varDelta'=0$ to reveal that Y_n is a background pixel, or $\mathrm{H}_0:\varDelta'=0$ versus $\mathrm{H}_1:\varDelta'=\varDelta$ to reveal that Y_n belongs to the object.

Put $\mathbf{\Delta} = (\varDelta,\varDelta')$ and let

$$G_n(x;\mathbf{\Delta}) = P_{\mathbf{\Delta}}[\widehat{m}_n \leqslant x], \quad x \in \mathbb{R},$$

denote the d.f. of the clipping median, where $P_{\mathbf{\Delta}}$ indicates that the probability is calculated assuming the mixture model given by (8)–(10) holds true.

Theorem 2. *Let $Y_1,\ldots,Y_n$ be independent random variables such that* (8)–(10) *hold true. Let $\mathscr{I}_1+\mathscr{I}_2=\{1,\ldots,n-1\}$ be a disjoint decomposition in two non-empty sets. For all $x \in R$ we have*

$$G_n^{(\mathscr{I}_1,\mathscr{I}_2)}(x;\mathbf{\Delta}) = \int \sum_{(k,i)\in\mathscr{J}(x,y)} P_{1,|\mathscr{I}_1|}(x,y,k)P_{2\varDelta,|\mathscr{I}_2|}(x,y,i)\,\mathrm{d}F(y-\varDelta'),$$

where $\mathscr{J}(x,y)=\{(k,i)\in\mathbb{N}^2 : |k|\leqslant|\mathscr{I}_1|, |i|\leqslant|\mathscr{I}_2|, k+i\leqslant 2\mathbf{1}(y>k(0)/M)\}$, *and*

$$P_{1,l}(x,y,r)=\sum_{i=-r}^{l-r} q(|\mathscr{I}_1|, i, r+i, p_+(x,y;0), p_-(x,y,0)), \quad |r|\leqslant|\mathscr{I}_1|, \tag{11}$$

$$P_{2\Delta,l}(x,y,r)=\sum_{i=-r}^{l-r} q(|\mathscr{I}_2|, i, r+i, p_+(x,y;\Delta), p_-(x,y;\Delta)), \quad |r|\leqslant|\mathscr{I}_2| \tag{12}$$

with

$$p_+(x,y;\delta)=P(|Y+\delta-y|\leqslant M, k([Y+\delta-y]/M)Y\geqslant x), \tag{13}$$

$$p_-(x,y;\delta)=P(|Y+\delta-y|\leqslant M, k([Y+\delta-y]/M)Y<x) \tag{14}$$

for $\delta\in\mathbb{R}$.

Proof. The proof is similar as the proof of Theorem 1. Let $\widetilde{Y}_1,\ldots,\widetilde{Y}_n \overset{\text{i.i.d.}}{\sim} F(x)$. Then we may assume $Y_i=\widetilde{Y}_i$ if $i\in\mathscr{I}_1$, $Y_i=\widetilde{Y}_i+\Delta$ if $i\in\mathscr{I}_2$, and $Y_n=\widetilde{Y}_n+\Delta'$. Using the same notation as in the proof of Theorem 1, we have given $Y_n=y$

$$L(y;\Delta)=\sum_{i\in\mathscr{I}_1}\mathbf{1}(|\widetilde{Y}_i-y|\leqslant M)+\sum_{i\in\mathscr{I}_2}\mathbf{1}(|\widetilde{Y}_i+\Delta-y|\leqslant M)+1,$$

$$\begin{aligned} S_{L(y)}(x;\Delta)=&\sum_{i\in\mathscr{I}_1}\mathbf{1}(|\widetilde{Y}_i-y|\leqslant M)\mathbf{1}(\widetilde{Z}_i(y;0)\geqslant x)\\ &+\sum_{i\in\mathscr{I}_2}\mathbf{1}(|\widetilde{Y}_i+\Delta-y|\leqslant M)\mathbf{1}(\widetilde{Z}_i(y;\Delta)\geqslant x)+\mathbf{1}(y\geqslant x/k(0)),\end{aligned}$$

where

$$\widetilde{Z}_i(y;\delta)=\mathbf{1}(k([\widetilde{Y}_i+\delta-y]/M)(\widetilde{Y}_i+\delta),$$

$i=1,\ldots,n$. Now $S_{L(y)}(x;\Delta)-L(y;\Delta)/2\leqslant 0$ is equivalent to

$$C(x,y;\Delta)\leqslant 2\mathbf{1}(y\geqslant k(0)/M), \tag{15}$$

where

$$C(x,y;\Delta)=S_1(x,y)+S_2(x,y;\Delta)$$

with

$$S_1(x,y)=\sum_{i\in\mathscr{I}_1}\eta_i(x,y;0), \quad S_2(x,y;\Delta)=\sum_{i\in\mathscr{I}_2}\eta_i(x,y;\Delta)$$

and

$$\eta_i(x,y;\delta)=2\mathbf{1}(|\widetilde{Y}_i+\delta-y|\leqslant M)\{\mathbf{1}(\widetilde{Z}_i(y;\delta)\geqslant x)-1/2\},$$

$i=1,\ldots,n-1$. Clearly,

$$S_1(x,y) \sim P_1(x,y;\circ) \quad \text{and} \quad S_2(x,y;\varDelta) \sim P_{2\varDelta}(x,y;\circ),$$

where the probability functions $P_1(x,y;\circ)$ and $P_{2\varDelta}(x,y;\circ)$ are defined in (11) and (12). By independence of $S_1(x,y)$ and $S_2(x,y;\varDelta)$, we have

$$P(S_1(x,y)+S_2(x,y) \leqslant 2\mathbf{1}(y>k(0)/M) \mid Y_n = y) = \sum_{(k,i)\in\mathscr{J}(x,y)} P_{1,|\mathscr{I}_1|}(x,y,k) P_{2\varDelta,|\mathscr{I}_2|}(x,y,i).$$

Therefore, we obtain

$$G_n^{(\mathscr{I}_1,\mathscr{I}_2)}(x;\mathbf{\Delta}) = \int \sum_{(k,i)\in\mathscr{J}(x,y)} P_{1,|\mathscr{I}_1|}(x,y,k) P_{2\varDelta,|\mathscr{I}_2|}(x,y,i)\,\mathrm{d}F(y-\varDelta'). \qquad \square$$

The definition of appropriate estimators strongly depends on the application. For instance, for certain applications in image processing one would prefer to estimate background and/or foreground using external data. If no external data is available or a local estimation procedure seems to be more appropriate one can proceed as follows. Calculate residuals

$$\widehat{\varepsilon}_i = Y_i, \quad i \in \mathscr{I}_1,$$
$$\widehat{\varepsilon}_i = Y_i - \widehat{\varDelta}, \quad i \in \mathscr{I}_2,$$

where, e.g. $\widehat{\varDelta} = |\mathscr{I}_2|^{-1}\sum_{i\in\mathscr{I}_2} Y_i$. Let $\widehat{F}_n(x) = (n-1)^{-1}\sum_{i=1}^{n-1}\mathbf{1}(\widehat{\varepsilon}_i \leqslant x)$. The d.f. of the clipping median under the mixture model can now be estimated by

$$\widehat{G}_{L_1,L_2,n}^{(\mathscr{I}_1,\mathscr{I}_2)}(x;\mathbf{\Delta}) = \int \sum_{(k,i)\in\mathscr{J}^*(x,y)} \widehat{P}_{1,L_1}(x,y,k)\widehat{P}_{2\varDelta,L_2}(x,y,i)\,\mathrm{d}\widehat{F}_n(y-\varDelta'),$$

where $\mathscr{J}^*(x,y) = \{(k,i)\in\mathbb{N}^2 : |k-(n-1)p_+(x,y)| \leqslant L_1, |i-(n-1)p_-(x,y)| \leqslant L_2, k+i \leqslant 2\mathbf{1}(y> k(0)/M)\}$ for truncation constants L_1 and L_2. Further,

$$\widehat{P}_{1,L_1}(x,y,r) = \sum_{i=-r}^{L_1-r} q(|\mathscr{I}_1|, i, r+i, \widehat{p}_{+n}(x,y;0), \widehat{p}_{-n}(x,y,0)), \quad |r| \leqslant |\mathscr{I}_1|,$$
$$\widehat{P}_{2\varDelta,L_2}(x,y,r) = \sum_{i=-r}^{L_2-r} q(|\mathscr{I}_2|, i, r+i, \widehat{p}_{+n}(x,y;\varDelta), \widehat{p}_{-n}(x,y;\varDelta)), \quad |r| \leqslant |\mathscr{I}_2|$$

with

$$\widehat{p}_{+n}(x,y;\delta) = (n-1)^{-1}\sum_{i=1}^{n-1}\mathbf{1}(|\widehat{\varepsilon}_i+\delta-y| \leqslant M, k([\widehat{\varepsilon}_i+\delta-y]/M)\widehat{\varepsilon}_i \geqslant x),$$
$$\widehat{p}_{-n}(x,y;\delta) = (n-1)^{-1}\sum_{i=1}^{n-1}\mathbf{1}(|\widehat{\varepsilon}_i+\delta-y| \leqslant M, k([\widehat{\varepsilon}_i+\delta-y]/M)\widehat{\varepsilon}_i < x)$$

for $\delta \in \mathbb{R}$.

Now we may test $\mathrm{H}_0 : \Delta' = \Delta$ versus $\mathrm{H}_1 : \Delta' \neq \Delta$ using the test

$$\phi_n = \mathbf{1}(\widehat{p}_{L_1,L_2,n} \notin [\alpha/2, 1-\alpha/2]). \tag{16}$$

where

$$\widehat{p}_{L_1,L_2,n} = \widehat{G}^{(\mathscr{I}_1,\mathscr{I}_2)}_{L_1,L_2,n}(\widehat{m}_n; \widehat{\mathbf{\Delta}})$$

is the estimated p-value.

4. Convergence of p-values

It remains to discuss the convergence of the proposed estimators for the p-values associated to the testing problem. We consider the general mixture model case studied in the previous section. The following theorem provides sufficient conditions for a.s. convergence of the distance between $\widehat{p}_{L_1,L_2,n}$ and the truncated p-value

$$G^{(\mathscr{I}_1,\mathscr{I}_2)}_{L_1,L_2,n}(\widehat{m}_n; \Delta, \Delta) = \int \sum_{(k,i)\in\mathscr{J}^*(\widehat{m_n,y})} P_{1,L_1}(\widehat{m}_n, y, k) P_{2\Delta,L_2}(\widehat{m}_n, y, i)\, \mathrm{d}F(y-\Delta).$$

For sufficiently large L_1 and L_2 the distance to the true p-value $G^{(\mathscr{I}_1,\mathscr{I}_2)}_n(\widehat{m}_n; \Delta, \Delta)$ is small.

Theorem 3. *Let F be a Lipschitz continuous d.f. Assume* (8)–(10), *and*

$$0 < c_1 \leqslant \frac{|\mathscr{I}_1|}{n}, \quad \frac{|\mathscr{I}_2|}{n} \leqslant c_2 < 1$$

for constants c_1, c_2. *Suppose*

$$\widehat{\Delta}_n \stackrel{\text{a.s.}}{\to} \Delta, \quad n \to \infty, \tag{17}$$

and

(*KC*) *The kernel k is continuous and the sets* $\{\widetilde{y} : k([\widetilde{y} - (y-\delta)]/M)\widetilde{y} \geqslant x\}$ *are intervals* $[A(x,y,\delta), B(x,y,\delta)]$ *with continuous functions A and B*.

Then

$$\sup_x |\widehat{G}^{(\mathscr{I}_1,\mathscr{I}_2)}_{L_1,L_2,n}(x; (\widehat{\Delta}_n, \widehat{\Delta}_n)) - G^{(\mathscr{I}_1,\mathscr{I}_2)}_{L_1,L_2,n}(x; (\widehat{\Delta}_n, \widehat{\Delta}_n))| \stackrel{\mathrm{P}}{\to} 0$$

as $n \to \infty$.

Remark 4.1. Condition (KC) is satisfied by many kernels, e.g., the Gaussian kernel for which $j(z) = zK([z-a]/M)$, $a \in \mathbb{R}$, is unimodal and concave. The implicit function theorem ensures that the solutions $z = z(c, a, M)$ of $j(z) = x$, $x > 0$, depend continuously on c, a, and M. Thus A and B are continuous.

Proof. Clearly, $\widehat{\Delta}_n \stackrel{\text{a.s.}}{\to} \Delta$, $n \to \infty$, implies that

$$\max_{i=1,\dots,n} |\widehat{\varepsilon}_i - \varepsilon_i| \stackrel{\text{a.s.}}{\to} 0$$

as $n \to \infty$, where $\varepsilon_i = Y_i$ if $i \in \mathscr{I}_1$, $\varepsilon_i = Y_i - \Delta$, if $i \in \mathscr{I}_2$, $\widehat{\varepsilon}_i = Y_i$, $i \in \mathscr{I}_1$, and $\widehat{\varepsilon}_i = Y_i - \widehat{\Delta}_n$, $i \in \mathscr{I}_2$. Therefore the e.d.f. of $\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_n$ converges to $F(x)$, uniformly in $x \in \mathbb{R}$. This implies weak convergence, i.e.,

$$\int g(y)\, \mathrm{d}\widehat{F}_n(y - \Delta') \to \int g(y)\, \mathrm{d}F(y - \Delta') \tag{18}$$

a.s., as $n \to \infty$, for all bounded and continuous functions g. Note that for $\mathbb{R}$-valued functions g_n, g defined on $\mathbb{R}^3$ the estimate

$$\begin{aligned} &\left| \int \widehat{g}_n(x,y,\delta)\, \mathrm{d}\widehat{F}_n(y) - \int g(x,y,\delta)\, \mathrm{d}F(y) \right| \\ &\quad \leqslant \sup_{x,y,\delta} |\widehat{g}_n(x,y,\delta) - g(x,y,\delta)| + \int g(x,y,\delta)\, \mathrm{d}(\widehat{F}_n - F)(y) \end{aligned}$$

also implies

$$\sup_{x,\delta} \left| \int \widehat{g}_n(x,y,\delta)\, \mathrm{d}\widehat{F}_n(y) - \int g(x,y,\delta)\, \mathrm{d}F(y) \right| \overset{\text{a.s.}}{\to} 0$$

as $n \to \infty$, if

$$\|\widehat{g}_n - g\|_\infty \overset{\text{a.s.}}{\to} 0 \tag{19}$$

as $n \to \infty$, and $\|g\|_\infty < \infty$.

Note that

$$\widehat{G}^{(\mathscr{I}_1,\mathscr{I}_2)}_{L_1,L_2,n}(x; \widehat{\boldsymbol{\Delta}}) = \int \sum_{(k,i) \in \mathscr{J}^*(x, y+\widehat{\Delta}_n)} \widehat{P}_{1,L_1}(x, y + \widehat{\Delta}_n) \widehat{P}_{2\widehat{\Delta}_n, L_2}(x, y + \widehat{\Delta}_n)\, \mathrm{d}F(y).$$

Thus, the assertion follows if we verify (19) for the functions

$$\begin{aligned} \widehat{g}_n(x,y,\delta) &= \sum_{(k,i) \in \mathscr{J}^*(x,y+\delta)} \widehat{P}_{1,L_1}(x, y+\delta) \widehat{P}_{2\delta,L_2}(x, y+\delta), \\ g(x,y,\delta) &= \sum_{(k,i) \in \mathscr{J}^*(x,y+\delta)} P_{1,L_1}(x, y+\delta, k) P_{2\delta,L_2}(x, y+\delta, i) \end{aligned}$$

since $\|g\|_\infty \leqslant 1$. We start by showing that $\widehat{p}_{+n}(x,y;\delta) \to p_+(x,y;\delta)$, as $n \to \infty$, a.s. Define

$$\widetilde{p}_{+n}(x,y;\delta) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbf{1}(\varepsilon_i \in [(y-\delta) - M, (y-\delta) + M], k([\varepsilon_i - \delta)/M]\varepsilon_i \geqslant x).$$

Note that by (KC) there exist functions $I_1(x,y,\delta)$ and $I_2(x,y,\delta)$ such that

$$\widetilde{p}_{+n}(x,y;\delta) = \frac{1}{n-1} \sum_{i=1}^{n-1} \mathbf{1}(\varepsilon_i \in [I_1(x,y,\delta), I_2(x,y,\delta)]).$$

Thus, by uniformity of the Glivenko–Cantelli Theorem over VC classes (see Blum, 1955; DeHardt, 1971; Shorack and Wellner, 1986, Chapter 26, Section 1, Theorem 1) we obtain

$$\sup_{x,y,\delta} |\widetilde{p}_{+n}(x,y,\delta) - p_+(x,y,\delta)| \overset{\text{a.s.}}{\to} 0$$

as $n \to \infty$. Note that

$$|\widehat{p}_{+n}(x,y,\delta) - \widetilde{p}_{+n}(x,y,\delta)| = (n-1)^{-1} \sum_{i=1}^{n-1} [\xi_i'(x,y,\delta) + \eta_i'(x,y,\delta)],$$

where

$$\xi_i'(x,y,\delta) = \mathbf{1}(\varepsilon_i \notin [I_1(x,y,\delta), I_2(x,y,\delta)], \widehat{\varepsilon}_i \in [I_1(x,y,\delta), I_2(x,y,\delta)]),$$
$$\zeta_i'(x,y,\delta) = \mathbf{1}(\varepsilon_i \in [I_1(x,y,\delta), I_2(x,y,\delta)], \widehat{\varepsilon}_i \notin [I_1(x,y,\delta), I_2(x,y,\delta)]),$$

$i = 1, \ldots, n$. Let $\eta > 0$. On a set Z with $P(Z) = 1$ there exists a $n_0 \in \mathbb{N}$ with $E_n = \max_{1 \leqslant i \leqslant n} |\widehat{\varepsilon}_i - \varepsilon_i| < \eta$ for all $n \geqslant n_0$. Writing $\widehat{\varepsilon}_i = \varepsilon_i + (\widehat{\varepsilon}_i - \varepsilon_i)$ we may estimate

$$\begin{aligned} \xi_i'(x,y,\delta) &\leqslant \mathbf{1}(\varepsilon_i \notin [I_1(x,y,\delta), I_2(x,y,\delta)], \varepsilon_i \in [I_1(x,y,\delta) - \eta, I_2(x,y,\delta) + \eta]) \\ &\leqslant \xi_i(x,y,\delta) = \mathbf{1}(\varepsilon_i \in [I_1(x,y,\delta) - \eta, I_1(x,y,\delta)] \cup [I_2(x,y,\delta), I_2(x,y,\delta) + \eta]) \end{aligned}$$

and, similarly,

$$\zeta_i'(x,y,\delta) \leqslant \zeta_i(x,y,\delta) = \mathbf{1}(\varepsilon_i \in [I_1(x,y,\delta), I_1(x,y,\delta) + \eta] \cup [I_2(x,y,\delta) - \eta, I_2(x,y,\delta)].$$

Again, since the Glivenko–Cantelli Theorem holds uniformly over VC classes,

$$\sup_{x,y,\delta} \left| (n-1)^{-1} \sum_{i=1}^{n-1} \xi_i(x,y,\delta) - P(\xi_1(x,y,\delta) = 1) \right| \overset{\text{a.s.}}{\to} 0$$

as $n \to \infty$. By Lipschitz continuity of F we have $P(\xi_i(x,y,\delta) = 1) = O(\eta)$ where the O does not depend on x, y, δ. Similarly, $\sup_{x,y,\delta} |(n-1)^{-1} \sum_{i=1}^{n-1} \zeta_i(x,y,\delta)| \overset{\text{a.s.}}{\to} O(\eta)$, as $n \to \infty$. Hence, we obtain

$$\sup_{x,y,\delta} |\widehat{p}_{+n}(x,y;\delta) - p_+(x,y;\delta)| \overset{\text{a.s.}}{\to} 0, \quad \sup_{x,y,\delta} |\widehat{p}_{-n}(x,y;\delta) - p_-(x,y;\delta)| \overset{\text{a.s.}}{\to} 0 \tag{20}$$

as $n \to \infty$. Note that this implies

$$q(i,j,\widehat{p}_{+n}(x,y;\delta), \widehat{p}_{-n}(x,y;\delta)) \overset{\text{a.s.}}{\to} q(i,j,p_+(x,y;\delta), p_-(x,y;\delta))$$

uniformly in (x,y,δ) and $|i| < I$ and $|j| < K$, as $n \to \infty$. Consequently, the difference of the corresponding d.f.s converges uniformly in (x,y,δ) to 0, i.e., for all $x', y' \in \mathbb{R}$,

$$\left| \sum_{k \leqslant x'} \sum_{l \leqslant y'} q(k,l,\widehat{p}_{+n}(x,y;\delta), \widehat{p}_{-n}(x,y;\delta)) - \sum_{k \leqslant x'} \sum_{l \leqslant y'} q(k,l,p_+(x,y;\delta), p_-(x,y;\delta)) \right| \to 0$$

as $n \to \infty$, w.p. 1, since there are only a finite number of summands. Noting that by definition $\widehat{P}_{1,L_1}$, P_{1,L_1}, $\widehat{P}_{2,\delta,L_2}$, and $P_{2\delta,L_2}$ are finite sums, (19) follows. $\square$

Table 1
Accuracy of the estimated null distribution for a two-sided significance test with nominal level $\alpha = 0.05$

M	h		
	25	50	100
0.5	0.0360	0.0360	0.0466
0.75	0.0372	0.0522	0.0554
1	0.0488	0.0512	0.0494
1.25	0.0418	0.0490	0.0526
1.5	0.0428	0.0478	0.0446
1.75	0.0352	0.0450	0.0448
2	0.0244	0.0360	0.0468

Table 2
Level and power of the test (16) to test the null hypothesis that the current pixel belongs to the object against the alternative that it belongs to the background for different sample sizes and parameter settings

Sample size h	background k	M			
		0.5	1	1.5	2
Size ($\Delta' = \Delta = 1$)					
24	6	0.041	0.021	0.021	0.025
	12	0.036	0.024	0.019	0.015
48	12	0.026	0.034	0.033	0.021
	24	0.024	0.017	0.028	0.018
63	24	0.027	0.023	0.041	0.040
	48	0.024	0.028	0.032	0.023
Power for $\Delta' = 0, \Delta = 1$					
24	6	0.097	0.118	0.121	0.079
	12	0.090	0.115	0.109	0.071
48	12	0.120	0.161	0.151	0.101
	24	0.126	0.161	0.144	0.106
63	24	0.144	0.149	0.167	0.116
	48	0.129	0.157	0.165	0.116

5. Simulations

We conducted a simulation study to assess the accuracy of the proposed methods. In our first experiment i.i.d. $N(0,1)$ samples were generated, and the clipping median $\widehat{m}_n$ using the Gaussian kernel was calculated. For the simulation we used estimators without truncation, but the resulting procedure is very time consuming. Therefore, for time critical applications we recommend to use the proposed truncated estimators. Table 1 reports the simulated level of the test which rejects the null hypothesis that $\Delta' = 0$ if the estimated p-value is less than 0.025 or greater 0.975. Each entry was obtained by 5000 repetitions. It can be seen that the estimated p-values provide reliable statistical tests.

The second experiment deals with the more interesting mixture model. Test samples were simulated according to the model

$$Y_i \sim \mathcal{N}(0,1), \quad i \in \mathcal{I}_1, \quad Y_i \sim \mathcal{N}(\Delta,1), \quad i \in \mathcal{I}_2$$

and $Y_n \sim \mathcal{N}(\Delta',1)$. We used a Gaussian kernel (standard normal density) for k.

To mimick an object detection problem, we used $h = 24, 48$, and 63, and $M = 0.5, 1, 1.5$, and 2 . The level shifts were chosen as $\Delta = 1$ and $\Delta' = 0, 1$. The number of background pixels k was chosen to ensure $k/h = 0.25$ and 0.5. Since the computations are more time consuming in the mixture model case, each table entry was calculated based on 1000 replications.

Table 2 provides estimates for the true level of the proposed procedure for a nominal level $\alpha = 0.05$ using $\Delta' = \Delta = 1$, and power estimates for $\Delta' = 0$ and $\Delta = 1$. It can be seen that the test procedure is conservative and also has power to detect small to moderate shifts. The results also indicate that there is an optimal value of M which maximizes the power.

Acknowledgements

The work of the author was supported by the Deutsche Forschungsgemeinschaft, SFB 475, Reduction of Complexity in Multivariate Data Structures.

References

Blum, J., 1955. On the convergence of empiric distributions. Ann. Math. Statist. 26, 527–529.

Budinger, T., et al., 1996. Mathematics and Physics of Emerging Biomedical Imaging. National Academic Press, Washington, DC.

Chiu, C.K., Glad, I.K., Godtliebsen, F., Marron, J.S., 1998. Edge-preserving smoothers for image processing. J. Amer. Statist. Assoc. 93, 526–556.

DeHardt, J., 1971. Generalizations of the Glivenko–Cantelli theorem. Ann. Math. Statist. 42, 2050–2055.

Godtliebsen, F., 1991. Noise reduction using Markov random fields. J. Magn. Reson. 92, 102–114.

Godtliebsen, F., Marron, S., 1997. A nonlinear Gaussian filter applied to images with discontinuities. J. Nonparametrics Statist. 8, 21–43.

Godtliebsen, F., Spjøtvoll, E., 1991. Comparison of statistical methods in MR imaging. Internat. J. Imag. Systems Tech. 3, 33–39.

Lee, J.S., 1983. Digital image smoothing and the sigma filter. Comput. Vision Graph. 24, 255–269.

Pawlak, M., Rafajłowicz, E., 2000. Vertically weighted regression—a tool for non-linear data analysis and constructing control charts. Amer. Statist. Assoc. 84, 367–388.

Pawlak, M., Rafajłowicz, E., 2001. Jump preserving signal reconstruction using vertical weighting. Nonlinear Anal—Theory 47, 327–338.

Pawlak, M., Rafajłowicz, E., Steland, A., 2004. Detecting jumps in time series—nonparametric setting. J. Nonparametrics Statist. 16, 329–347.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Steland, A., 2002a. Nonparametric monitoring of financial time series by jump-preserving estimators. Statist. Papers 43, 361–377.

Steland, A., 2004a. Jump-preserving monitoring of dependent time series under local alternatives. Statist. Decisions 21 (4), 343–366.

Steland, A., 2004b. Sequential control of time series by functionals of kernel-weighted empirical processes under local alternatives. Metrika, 60, 229–249.

Wolfram Research, 2004. Application documentation: digital image processing package, Section 5.5 ⟨http://documents.wolfram.com/applications/digitalimage⟩.